Diagnosis: from classification to prediction

David Armstrong

King's College London

Department of Primary Care and Public Health Sciences

Addison House

Guy's Campus

London SE1 1UL

david.armstrong@kcl.ac.uk

Diagnosis: from classification to prediction

Abstract:

Over the last two decades diagnostic labels have increasingly been sub-divided based on molecular and genetic 'signatures'. But this emphasis on disease sub-types defined in molecular terms, elides the central role of population-based predictive technologies in determining these new diagnoses. While molecular diagnostic sub-types might flow from the laboratory, the clinical validity of every putative diagnostic category must ultimately be tested against its predictive powers. In effect, the former logic of prognosis following diagnosis is reversed. This paper explores the emergence of this new method of diagnostic practice over the last half century.

Keywords: diagnosis, classification, prognosis, prediction

Words: main body 5,219, total 6,652

Cancer of the breast is a long-established clinical diagnosis. The malignant tumour increases in size, spreads to lymph nodes and then to the rest of the body. This singular diagnosis was challenged in 1967 when it was suggested there were in fact two types of breast cancer (Hems, 1967). A decade or so later the World Health Organisation (International histological classification of tumours, 1981) argued there were 14 types. A more recent review of breast cancer identified over 20 different 'types' supporting the contention that breast cancer was "a complex and heterogeneous disease, comprising multiple tumor entities associated with distinctive histological patterns and different biological features and clinical behaviors" (Weigelt & Reis-Filho, 2009, p. 718). In short, breast cancer was not "a single disease, but is instead a collection of diseases that have distinct histopathological features, genetic and genomic variability, and diverse prognostic outcomes" (Vargo-Gogola & Rosen, 2007, p. 659).

The term 'diagnostic heterogeneity' was introduced in the 1980s to describe the procession of sub-dividing diagnostic categories, a descriptor that captured the plurality of 'conditions' underlying any diagnosis. Recognition that heterogeneity characterised traditional unitary diagnoses marked a new way of constituting a disease label. Diseases had existed in a classification space or table with individual 'cells' (diagnoses) being added, subtracted or amended. Sometimes new diagnoses resulted from the discovery of new pathological entities. Sometimes diagnoses were merged when it was determined that they had the same underlying pathology as occurred when phthisis, scrofula and consumption were found to be all manifestations of tuberculosis. And sometimes diagnoses were split as in the identification of Type 1 and Type II diabetes early in the 20th century. All these diagnostic

innovations were the result of a more penetrating analysis of pathology, of the nature and causes of the biological processes underpinning the disease in question. What was new in the late 20th century concept of diagnostic heterogeneity, of disease segmentation, was the role of prediction in drawing the boundaries between new disease categories. While the classification of disease had been based on pathology, the new system for generating disease categories was constructed with reference to the future.

In a study of medical textbooks published between 1892 and 1988, Christakis (1997) noted the gradual disappearance of concerns with and reference to prognosis, that is the future course of the disease diagnosed by the doctor. In part, he explained this 'ellipsis' in terms of an important shift in the ontological status of disease. Whereas in the 19th century the prognosis was highly dependent on the patient's unique characteristics, especially their constitution, in the early 20th century it was believed that two patients could have the same disease whose natural course was a function of its pathological nature (Rosenberg, 2002). The prognosis of lobar pneumonia, for example (the medical focus of Christakis' paper), was subsumed within the diagnosis itself and its 'typical' natural history. This process was promoted, according to Christakis, by the advent of antibiotics, an effective treatment for pneumonia, that further diminished the importance of considering prognosis as a separate analytic category.

While prognosis might have almost disappeared from clinical thought and practice, there is evidence for its recent resurgence in the late 20th century as indicated by PubMed citations. Bourret, Keating and Cambrosio (2011) have argued that (clinical) prognosis and prediction

have distinct meanings with the former referring to 'will the disease recur?' and the latter 'will the disease react to treatment?', this distinction being ascribed to 'the (recent) emergence of molecular technologies' (p.2) – though elsewhere they described this distinction as 'often clouded' in practice (Kohli-Laven, Bourret, Keating, & Cambrosio, 2011). The argument here, however, is that both terms now have a common basis derived not from molecular technologies but from patient populations and have therefore come to be used interchangeably. In other words, the idea of prognosis as a property of the pathology and its natural history, as described by Christakis, has been replaced by a future orientation. The latter is predicated on the late 20th century idea of a population in which the earlier enumeration of individuals has been superseded by the multiple and virtual populations that can be inferred from any patient characteristic (Armstrong, 2017). These (sub) populations provided the essential denominator for calculating differential prognoses.

The argument for the claim that population-based prediction has begun to replace the role of pathology (and more recent molecular and genetic developments) in diagnostic practice is presented in three case studies. These are based on primary sources so that the 'conditions of possibility' for emergence might be described rather than the social, political and economic explanations favoured by historians. Indeed, a logic of prediction may just as well be used to explain many of those same social, political and economic 'influences'. Whether, say, the activities of the pharmaceutical industry are responsible for these changes or whether they simply exploit them is difficult to determine and is a discussion not pursued in this paper. The analysis here is an attempt at writing a genealogy of diagnosis, a

history of the present (Garland, 2014), that seeks to uncover the implicit logic sustaining diagnostic formation in recent decades.

The case studies are mainly based on searches of the medical literature (using resources such as JSTOR and PubMed, albeit restricted to English language publications) with reference to contemporary sociological literature where appropriate. Each case study – the boundary problem of hypertension, the process of staging cancers and the development of severity indices – provides an instance of when prediction began to undermine the older epistemology of pathological classification. While the latter two cases are marked by the beginning of a new process of diagnostic fragmentation, the first example concerns the struggle to maintain a distinct boundary around the pathology of hypertension.

The boundaries of hypertension

Until the end of the 19th century high blood pressure (hypertension) was diagnosed by placing a finger on the pulse at the wrist to 'feel' the tension. But with the invention of the sphygmomanometer it became possible to measure more precisely the underlying (diastolic) pressure in the patient's arteries, together with the (systolic) pressure pulse caused by the beating heart. If the pressure was too high it could cause blood to burst out of an artery and haemorrhage into surrounding organs and tissues. The diagnosis of hypertension was complicated by the observation that blood pressure tended to rise with age so that defining a normal range proved challenging.

In 1926 US insurance companies pooled the blood pressure readings from 700,000 applicants for life insurance to try and establish a more precise range of 'normal' pressures for different ages (Hunter, 1926). Insured populations, however, depended on who applied for life insurance and who was accepted; and they included those with pre-existing hypertension and those without. Insurance data was therefore "distorted" as no attempt was made "to separate normal from abnormal blood pressure groups in the statistical treatment" (Robinson & Brucer, 1939, p. 413). The problem was a circular one. To be able to identify the 'normal' range of blood pressures in a population the abnormal ones had first to be removed but the latter could only be identified once the normal range was defined.

The challenge of defining the normal range of blood pressure was further explored in two major population-based initiatives in the second half of the 20th century, the Framingham Heart and the Build and Blood Pressure Studies. The Framingham Heart Study involved monitoring a population in a Massachusetts' town to determine which levels of blood pressure would prove harmful in terms of subsequent cardiovascular disease. The first patient was recruited in 1948 and by 1957, with early results emerging, the investigators noted a nearly four-times increase in incidence of coronary heart disease for study participants with hypertension, defined as a blood pressure of 160/95 mm Hg or higher (Dawber, Moore, & Mann, 1957). A few years later, however, the Framingham investigators reported that 'elevation of blood pressure' – yet still within the normal range – was associated with an increased risk of the development of heart disease in certain age groups (Kannel, Dawber, Kagan, Revotskie, & Stokes, 1961, p. 40). This 'elevated' blood pressure was labelled 'borderline hypertension', a term first proposed in an earlier study that

suggested hypertension might be defined statistically as outside two standard deviations of the mean (Master, Dublin, & Marks, 1951). Yet while the label of borderline hypertension seemed to indicate some increased risk of heart disease, the report was clear that "Normal means not abnormal; it includes normotensive and borderline hypertension" (Kannel *et al*., 1961, p. 44).

Similar findings were reported by the Build and Blood Pressure Study (1959) of the Society of Actuaries based on a population of 3,900,000 insurance policy holders and 102,000 deaths. It showed that mortality ratios were more markedly affected by blood pressure than earlier studies had indicated and that the risks extended into the accepted normal range. The analysis was still framed in terms of pathology: high blood pressure or hypertension could still be differentiated from normal blood pressure, the only question was exactly where that boundary should be drawn. And whereas the Framingham investigators had used the term borderline hypertension, the Build and Blood Pressure study suggested 'mild hypertension' to describe this indeterminate region between the pathological and the normal.

So long as hypertension was viewed as a disease, as something pathological to be contrasted with a healthy 'normal' blood pressure, the boundary continued to be problematic. "Many terms have been proposed to describe borderline hypertension. Generally, the definition includes some discrimination between the severity ('borderline', 'mild', 'benign') and the stability of the blood pressure ('labile', 'occasional', 'transient') … Confusing nomenclature and various definitions seriously interfere with meaningful

comparison of results of different investigators" (Julius & Schork, 1971, p. 746). What exactly was borderline hypertension? Was it even a specific entity? (Irving, Brash, Kerr, & Kirby, 1975). Should it be treated? (Lancet editorial, 1975). It was estimated that 18 million Americans had 'marginal elevations in blood pressure' and their management was 'a perplexing problem' (Julius, 1977). Patients in this no-man's land between normality and disease existed where guidelines for intervention were not established given that the "specific boundaries between segments designated as severe, moderate, mild, or otherwise are more accidental and arbitrary in nature" (Labarthe, 1978, p. 12).

By the late 1960s, the Framingham investigators were becoming more ambivalent about the usefulness of the hypertensive diagnostic label. They noted that knowledge of hypertension was based on clinical data derived from experience with clinical hypertensive disease, but their community study was largely based on early asymptomatic disease detected in the general population for which no guidelines existed. "An examination of the distribution of blood pressure in the general population fails to reveal any basis for selecting some critical value as the boundary between 'normotension' and 'hypertension' … the designation of some particular blood pressure value as 'hypertension' must of necessity be somewhat arbitrary" (Kannel, Castelli, McNamara, & Sorlie, 1969, pp. 118-9). This was the point at which hypertension, or rather the non-evaluative 'blood pressure', started to become a risk factor rather than a disease, a continuous variable located in a population of patients whose value was not underpinned by pathology but by its clinical consequences. Recognition that blood pressure need not be categorised as 'hypertensive' or 'high' or 'raised' but rather treated as a continuous scale then enabled its incorporation into linear coefficients that

aggregated several risk factors as predictive markers of cardiovascular disease or mortality (Truett, Cornfield, & Kannel, 1967).

The boundary between pathology and normality continued to be a focus of medical enquiry whether in terms of borderline or mild hypertension or, more recently, prehypertension (Moser, 2004). Successive clinical trials gradually lowered the blood pressure levels at which health benefits of treatment could be detected. The SPRINT trial (SPRINT Research Group, 2015) of intensive blood pressure control, for example, claimed that reducing the systolic pressure to less than 120mg Hg – well within the historical 'normal range' – conferred health benefits. Clinical guidelines identified a new Stage 1 hypertension in place of the old prehypertension which, controversially, increased the proportion of US adults who might benefit from treatment from 32% to 46% (Bakris & Sorrentino, 2018). These shifts in blood pressure treatment advice indicated a convergence of concepts of risk and disease as former borderline conditions became eligible for treatment (Aronowitz, 2009). Indeed, the consequences of equating at risk status with illness has important consequences for turning preventive activity into treatment and the subsequent medicalisation of previously 'healthy' patients (Kreiner & Hunt, 2014). But while there are debates about whether a 'disease' or a 'risk factor' is being over-diagnosed or over-treated (Moynihan, 2012; Martin, Boucher, Wright, & Saini, 2014) and whether blood pressure should be dichotomised or treated as a continuous variable in risk assessment (Will, 2005), the underlying logic was that a diagnostic label was being defined by its downstream clinical effects.

High blood pressure was one of the first clinical diagnoses to illustrate the new approach to classification. The significance of a 'blood pressure' label lay not in its anatomical or physiological characteristics and even less in any underlying pathological process – in a way these were irrelevant to its identification – but in future risks that established a continuous scale of potential harms. The very idea of a risk factor had emerged from the Framingham cohort (Kannel *et al*, 1961) as risk could only be derived from a population rather than from an individual patient. The risk coefficients that have subsequently been developed have inverted the usual temporal ordering of diagnosis assuming and determining prognosis. Instead, the 'prognosis', in this case the mortality resulting from any risk factor, could be used to validate the significance of that risk factor. Blood pressure provided a model for how types, stages and levels of 'disease' could be calibrated against future health rather than being defined in terms of a pathological model of disease that existed in opposition to the healthy/normal.

Cancer staging

The concept of staging had been applied in the earlier part of the 20th century to surgical operations which used two step interventions to reduce post-surgery mortality. But in the second half of the century staging became more commonly used to describe the penetration of cancer into the patient's organs and tissues. It was apparent that cancer started from a small number of cells then spread (metastasised) around the body. Identifying how far the cancer cells had spread was important for determining the extent of the disease progression, and therefore likely prognosis, as well as indicating the appropriate type of treatment (medical, surgical, radiological or palliative).

A number of methods for staging cancers were developed in the immediate post-War years (Craver, 1947; Peters, 1950) but the most influential system was based on the extent of the primary tumour (T), the state of the regional lymph nodes (N), and the presence or absence of distant metastases (M) (Harmer, 1964). During the 1950s this classification system became widely recognised and over the following decade its reach extended to cancers at many different sites (Sellars, 1966). In all cases, the TNM score indicated the extent the cancer had grown and spread as well as giving a guide as to the most appropriate treatment (surgery, for example, was too late for tumours that had metastasised) (Jelliffe & Thomson, 1955).

Once one patient's cancer stage could be compared with another's, clinicians could communicate more clearly and begin to align their treatments to the defined stage of the disease. Different treatments could also be better evaluated by ensuring comparisons were based on standardised assessment of the disease phase (Newall, 1967). More significantly, the new staging system in cancer began a process of segmenting diagnoses. For cancers of the breast, for example, each component – T, N, and M – could be sub-divided into gradations, according to its stage of progression. Thus, T1, T2, T3, and T4 represented degrees of advance in the primary tumour, Nb, Nc, and Nd represented degrees of advance in the local lymph nodes while the extent of distant metastases could be represented by M1, M2, M3, and M4. A patient might have 'breast cancer' but the staging diagnosis meant there might be little equivalence with another patient with the same disease label.

The initial process of cancer staging was based on a morphological/pathological classification in which the cancer was usually designated histologically as a cellular type, and then examined for whether it was anatomically localized, regionally extensive or distantly disseminated. The correspondence with outcome was not always accurate as some early stage cancers could progress rapidly while those at a later stage might develop slowly. The solution, as with the problem of blood pressure, was to start reversing the relationship between disease and its outcome. Instead of a stage being based on analysis of the spread of malignant cells, putative stages and cancer sub-types could be validated by their ability to predict outcome. Certainly, the idea of a cancer stage had implied a likely prognosis but the heterogeneity in outcomes of any particular stage could only be resolved by recalibrating the stage on the basis of adjustment from directly measured outcomes – leading to the derivation of further sub-types.

For some cancers a staging 'scale' was reconstructed and new variables were added to the model especially if these had a significant bearing on prognosis. The Gleason score for prostate cancer, for example, was first developed in the 1960s and 1970s by classifying cells by their pathological 'stage', from low-grade/normal to high-grade/aggressive. It was revised in 2005 and again in 2014, not on the basis of new investigations in pathology or the molecular laboratory nor on observations in clinical practice, but on its predictive value in a population. The 2014 revision, for instance, was tested and validated against 20,000 prostatectomy specimens and at least 16,000 biopsy samples to produce a 5-point Gleason Grade grouping that denoted a prognostically distinct stratification, from number 1 indicating lowest-risk cancer to 5 the most aggressive disease (Epstein *et al*., 2016). Similar

revisions were carried out for colon cancer and cervical cancer. Indeed, there were 14 references to cancer scoring systems in PubMed prior to 1980 but over 1,600 in the first decade of the following century.

The spread of screening services brought the early stages of cancer growth into greater focus. Many patients were diagnosed with 'carcinoma *in situ*', seemingly non-invasive cells undergoing changes that might lead to cancer. For the pathologist these changes were often subtle and difficult to distinguish from actively growing tissue (BMJ editorial, 1955). Many patients undergoing breast or cervical cancer screening, for example, were found to have 'cell dysplasia', some indeterminate changes from normal, or 'carcinoma *in situ*' characterised by possible cancer cells that had not spread. The CIN (Cervical Intraepithelial Neoplasia) nomenclature in cervical cancer screening placed all pre-malignant abnormalities into a single diagnostic scoring system ranging from mild dysplasia (CIN 1) to 'carcinoma-in-situ' (CIN 3). Even then there were calls to recognise 'borderline CIN' when CIN 1 was uncertain (Fox & Buckley, 1990).

'Borderline' findings might indicate a form of pre-cancer but how should it be treated? Would 'abnormal' cells grow further, and would their early removal prevent later invasive disease? The question could not be settled by histology but required examining populations of women to determine the significance of traditional staging, tumour biology and effect of treatment on prognosis (Saadatmand, Bretveld, Siesling, & Tilanus-Linthors, 2015). If there was debate about the status of *in situ* cancer it was only because the findings from population studies were uncertain or contradictory.

Initially, cancer staging had fragmented cancer diagnoses into sub-types based on the natural history of the underlying pathology but with the introduction of prognosis as a validating method, these sub-types began to proliferate to reflect the divergent futures of cancers even at the same TNM stage. In effect, as with hypertension, and at about the same time, cancer began to be categorised in terms of its likely prognosis. A similar process of initial diagnostic fragmentation followed by the application of a prognostic analytic framework also characterised other non-cancer diagnoses where the initial concern was with disease severity.

Severity indices

About the middle of the 20th century it became apparent that claims for the benefits of treatment were often based on invalid comparisons. Some treatments seemed better than others simply because they were used in less severe cases. In diabetic coma, for example, the patient's state of consciousness and age seemed to affect mortality rates so that the effect of a therapy on two groups of patients could not easily be compared. The solution was a Severity Index that "will permit comparison of experiences of one clinic with those of another with a reasonable degree of accuracy and thus afford a means of evaluating different methods of treatment in the future" (Rabinowitch, Fowler, & Bensley, 1939, p. 1424). Calculation of the diabetic coma Severity Index involved scoring on a five-point scale those variables known to affect mortality and adding them together. A few years later Collen (1942) added other variables that were known to affect mortality (such as the duration of the coma) to the Severity Index to improve its accuracy. Later still, Zieve and Hill

(1953) more formally analysed the discriminative power of each severity variable and, weighting each accordingly, showed the resulting single severity score was more accurate at predicting mortality than earlier simple summations.

The construction of severity indices multiplied in the second half of the 20th century with the aim of ensuring comparable groups of patients could be created for evaluating interventions. A severity index for hypertension could provide a bench-mark for judging whether a treatment was successful or not (Palmer, Loofbourow, & Doering, 1948); a 'Pathologic Index Rating' for patients with myocardial infarction enabled the effect of treatment to be more reliably assessed as patients with disease of similar severity could be compared (Schnur, 1953); a severity index for patients with hyaline-membrane disease could "facilitate evaluation of modifications of or additions to current regimens of treatment" (Stahlman, Battersby, Shepard, & Blankenship, 1967, p. 306); or for patients in shock after attempting suicide by drug overdose an index could be used "with particular reference to the effect of specific therapeutic interventions" (Afifi, Sacks, Liu, Weil, & Shubin, 1971, p. 502). Even in the face of increasing use of randomisation to generate comparable groups (Cochrane, 1972), severity indices were still held to have value. The Coronary Prognostic Index, for example, "should provide a valuable method for assessing randomisation in any properly designed trial which is aimed at comparing the progress of treated patients with those in a control group" (Norris, Brandt, Caughey, Lee, & Scott, 1969, p. 278). Severity scores, moreover, could be used for stratification to augment the randomisation process.

While the main purpose of measures of disease severity was to enable the assembly of equivalent patient groups for testing the value of interventions, severity was clearly related to prognosis as less severe or less advanced disease had a longer course to run. Sometimes, therefore, a severity index was described as a prognostic index. But at first there seemed little realisation that the analysis could easily be inverted: instead of severity indicating prognosis, prognosis could validate severity. That inversion began as a new focus on predictive power emerged. Previously "adjectives and adverbs rather than quantitative data" had underpinned prognosis; now "With the recent publication of mathematical expressions for the likelihood that a patient with a myocardial infarction will survive acutely and chronically, medicine took a large step forward in replacing verbal with quantitative description" (Bleich, 1971, p. 1533). It was the emerging indices and scoring systems that marked out the new diagnostic method: if the number assigned a disease or disease sub-type was based on its prognosis then that number was, almost by definition, a prognostic indicator. This number could then, working in reverse, validate a new disease sub-type or risk factor. Diagnosis therefore began to shift from an examination of pathologies (mostly by cellular though later by molecular techniques) to an investigation of prognoses based on predictive population technologies.

Although risk factors, cancer staging and severity indices had all originated in a conceptual universe dominated by pathology and pathological diagnoses, by the closing decades of the 20th century all three areas had been reoriented towards prognosis. The former discrete diagnostic labels based on an underlying 'disease' began a process of fragmentation that reflected smaller increments of risk. In this way, some diseases shifted towards becoming a

continuous scale – a blood pressure, a cancer stage, grade or score, a severity index – on which patients could be located with each step signifying a worsening prognosis. At the same time, the scale began to extend to incorporate more and more 'normal' patients. And because for any disease there were many possible futures, the potential for undermining the distinct pathological diagnostic categories of the present was considerable.

Diagnostic heterogeneity

The term 'translational research' emerged in the early 1990s to describe the process of moving basic laboratory research into the clinic. In the case of diagnosis, this involved the revitalisation of an older tradition of pathology with identification of molecular and genetic 'signatures' underpinning disease sub-types. Yet, ironically, despite diagnostic sub-divisions being expressed in molecular and genetic terms, they were in fact rooted in a completely different technology, namely the predictive power that derived from using population denominators. Pathology and the molecular laboratory produced 'candidate' molecular diagnoses, but they only entered clinical practice if they could be validated by predicting an endpoint in a clinical population. Indeed, the recent history of medicine is littered with thousands of molecular markers of disease that have failed to prove of any predictive value (Diamandis, 2012; Ioannidis & Bossuyt, 2017; Ioannidis & Tzoulaki, 2011; Kern, 2012). An extended classification system that, for example, suggested breast cancer might be sub-divided by gene expression profiling could only become part of the oncologist's diagnostic repertoire if the new sub-types had implications for the future course of the disease. But most laboratory speculations ended in failure: "Thousands of studies of putative prognostic or predictive markers have failed to identify useful molecular markers in tumors [because

they could not] predict the outcome in an individual patient with cancer or the response of an individual tumor to specific therapies" (Sauter & Simon, 2002, p. 1995).

The apparent molecular trend in recent diagnostic heterogeneity has therefore concealed the importance of a population-based predictive context. To be sure, molecular and genetic scientists/pathologists have generated innumerable diagnostic sub-divisions; theirs is a machine for fragmenting former unitary diagnoses as gross pathology and cellular histology is divided and sub-divided. But each one of these new putative diagnoses cannot enter the clinic without first passing through the evaluation gateway of the population-based predictive challenge (often in the form of a cohort study, a diagnostic accuracy test or a clinical trial). If a new genetic signature has no clinical significance in terms of predicting life expectancy, the clinical course of the disease or the most appropriate treatment then it failed to 'leave' the laboratory. Despite the apparent origin of stratified or precision medicine in the molecular laboratory, it is only realised through its accuracy in predicting outcomes in clinical populations.

Whereas in the past, treatment could influence prognosis, in the new regime prognosis (of a diagnostic sub-type) indicated treatment. Some patients responded to a treatment, others did not. Patients could therefore be sub-divided by their 'treatment responsiveness'. "Breast cancer is not a single disease but a group of several important tumor subtypes, each with a different natural history and each requiring a different treatment" (Burstein, 2005, p. 1652). A new type of diagnosis could then emerge: 'treatment resistant' disease, based not on pathology but on whether patients were 'responders' or 'non-responders' to treatment.

Drug-resistant syphilis, tuberculosis or depression, antimicrobial resistant infections, refractory 'difficult-to-control' hypertension or resistance to cancer chemotherapy were all identified through working backwards, from the outcome to the diagnosis.

The decline and rebirth of prognosis

In some ways prognosis is a term that belongs to a declining medical logic (Christakis, 1997), one that projects forward from the pathology, from the diagnostic label, with all the uncertainty that entails. The more recent approach, however, is perhaps better described as being underpinned by reverse prediction in which the outcome determines the diagnosis. This shift not only challenges the medical epistemological landscape but also changes clinical practice and begins to undermine a pathology-based diagnostic classification system.

Overall, there have been three major new approaches to diagnosis that were all underpinned by a temporal dimension. The first was prognosis/prediction. Different patients with the same disease often had different outcomes: some died, some lived, some recovered quickly, some diseases became chronic. Diagnosis could therefore be refashioned around these outcomes so that any illness categorisation more accurately reflected that predicted endpoint. Whether these 'diagnoses' were framed as risk factors or disease sub-types (Aronowitz, 2009), the effect was little different: each was predicated on a future trajectory. Given the range of possible outcomes this implied that the singular pathological diagnosis could fragment into increasing numbers of sub-types. Secondly, the presence of risk factors could be confirmed by their association with outcomes and the resulting

individual variables could further be clustered into prognostic indices or syndromes. Such risk factors, unlike pathological processes, had no necessary causal relationship with illness rather, like biomarkers, it was their role in prediction that validated their clinical utility. Finally, it was known that the response to treatment varied considerably between different patients. Treatment-resistant or treatment-recalcitrant disease therefore became an important sub-type of many diagnostic labels.

These three new diagnostic principles – prediction, risk factor clustering and treatment-responsiveness – have offered an alternative to the old classification systems over the last few decades. The flat diagnostic classificatory table could be transformed by introducing a temporal dimension in terms of the preceding risk factor of the disease and its likely temporal course with and without treatment. A diagnosis was no longer therefore a cell in a (pathological) classification table but an assignment to a 'population' of patients with similar prognostic trajectories. If such a population could not be identified then prognosis might remain unstated (Timmermans & Stivers, 2018).

For a system of clinical medicine that had relied on clinical skills to classify pathologies these new techniques were alien, as they were based on studies that identified population, sub-populations and sub-sub-populations. The fragmentation of formerly unitary diagnoses was not therefore a continuation of a clinical method for identifying new diseases but rather a different way of constructing diagnoses. A patient might be presented with a diagnosis of, say, breast cancer but the important question is 'what type?'. Some types are aggressive and deadly; in other cases the patient is likely to die with the cancer rather than from it;

some types are susceptible to the latest monoclonal antibodies while others show no response. Even these types have little stability as with more longitudinal studies and more therapeutic evaluations, sub-types will continue to change and proliferate. The very language of medicine reflects these changes. Severity and prognostic indices, staging, disease sub-types and diagnostic heterogeneity are all terms belonging to the period of transition as a new temporal future emerged that provided the techniques for illness identification in the present.

Over the last decade the sociology of diagnosis has come into increasing prominence (Jutel, 2009; McGann & Hutson, 2011; Jutel & Nettleton, 2011). “Diagnoses are the classification tools of medicine” claimed Jutel (2009, p. 278): “Diagnosis defines the field of medicine and its professional reach, serves as the nexus in which the clinical encounter takes place, arbitrates normality and difference, organises a patient’s illness, and determines how resources are allocated” (p. 294). According to this schema, diagnosis places patients in a classificatory frame that then determined treatment and prognosis. Yet, as argued above, while this classification frame has recently been sub-divided and segmented seemingly on the basis, as of old, of pathology, albeit increasingly molecular, this new apparently reductionist classificatory process is in fact premised on contexts far removed from the molecular laboratory. It is this new focus on the future as engendered by a population-based predictive technology that increasingly underpins the process of diagnosis.

References

Afifi, A.A., Sacks, S.T., Liu, V.Y., Weil, M.H., & Shubin, H. (1971). Accumulative prognostic index for patients with barbiturate, glutethimide and meprobamate intoxication. *New England Journal of Medicine, 285*(27), 1497-1502.

Armstrong, D. (2017). Clinical prediction and the idea of a population. *Social Studies of Science, 47*(2), 288 –299.

Aronowitz, R.A. (2009). The converged experience of risk and disease. *Milbank Quarterly, 87*(2), 417–442.

Bakris, G., & Sorrentino, M. (2018). Redefining Hypertension — Assessing the new blood-pressure guidelines. *New England Journal of Medicine, 378*(6), 497-9.

Bleich, H.L. (1971). Prognosis by calculation. *New England Journal of Medicine, 285*(27), 1533-4.

BMJ editorial. (1955). Carcinoma in situ. *British Medical Journal, 2*, 1609-1610.

Brown, P. (1990). The name game: Toward a sociology of diagnosis. *Journal of Mind and Behaviour, 11*(3-4), 385–406.

Bourret, P., Keating, P., & Cambrosio, A. (2011). Regulating diagnosis in post-genomic medicine: Re-aligning clinical judgment? *Social Science & Medicine*, *73*, 816-824.

Build and Blood Pressure Study. (1959). Chicago: Society of Actuaries.

Burstein, H.J. (2005). The distinctive nature of HER2-positive breast cancers. *New England Journal of Medicine, 353*(16), 1652-4.

Christakis, N.A. (1997). The ellipsis of prognosis in modern medical thought. *Social Science and Medicine, 44*(3), 301-315.

Cochrane, A.L. (1972). *Effectiveness and efficiency: Random reflections on health services*. London: Nuffield Trust.

Collen, M.F. (1942). Mortality in diabetic come. *Archives of Internal Medicine, 70*(3), 347-368.

Craver, L.F. (1947). Lymphomas and leukemias. *Bulletin of the New York Academy of Medicine, 23*(2), 79-100.

Dawber, T.R., Moore, F.E., & Mann G.V. (1957). Coronary heart disease in the Framingham study. *American Journal of Public Health, 47*, 4–24.

Diamandis, EP. (2012) The failure of protein cancer biomarkers to reach the clinic: why, and what can be done to address the problem? *BMC Medicine*, 10:87.

Epstein, J.I., Egevad, L., Amin, M.B., Delahunt, B., Srigley, J.R., Humphrey, P.A., & Grading Committee. (2016). The 2014 International Society of Urological Pathology (ISUP) Consensus Conference on Gleason Grading of Prostatic Carcinoma. *American Journal of Surgical Pathology, 40*(2), 244-252.

Fox, H., & Buckley, C.H. (1990). Current problems in the pathology of intra-epithelial lesions of the uterine cervix. *Histopathology, 17*(1), 1-6.

Garland, D. (2014). What is a ‘‘history of the present’’? On Foucault’s genealogies and their critical preconditions. *Punishment & Society, 16*(4), 365–384.

Harmer, M.H. (1964). Recent Advances in T.N.M. *British Medical Journal, 2*, 1319-1321.

Hems, G. (1967). Two types of breast cancer? *British Medical Journal, 2*, 496.

Hunter, A. (1926). Experience of American companies in regard to blood pressure. *Journal of the Institute of Actuaries, 57*(1), 35-44.

International histological classification of tumours No. 2. *Histological typing of breast tumours. (*1981). Second Edition. Geneva: WHO.

Ioannidis, J.P.A., & Bossuyt, P.M.M. (2017). Waste, leaks, and failures in the biomarker pipeline. *Clinical Chemistry, 63*(5), 963–972.

Ioannidis, J.P.A., & Tzoulaki, I. (2012). Minimal and null predictive effects for the most popular blood biomarkers of cardiovascular disease. *Circulation Research, 110*(5), 658-662.

Irving F.B., Brash H., Kerr, F., & Kirby, B.J. (1975). Proceedings: Borderline hypertension - a specific entity? *British Heart Journal, 37*(5), 549-50.

Jelliffe, A.M., & Thomson, D. (1955). The Prognosis in Hodgkin's Disease. *British Journal of Cancer, 9*(1), 21–36.

Julius, S. (1977). Borderline hypertension: An overview. *Medical Clinics of North America, 61*(3), 495-511.

Julius, S., & Schork, M.A. (1971). Borderline hypertension – A critical review. *Journal of Chronic Diseases, 23*(10-11), 723-754.

Jutel, A. (2009). Sociology of diagnosis: A preliminary review. *Sociology of Health & Illness, 31*(2), 278–299.

Jutel, A., & Nettleton, S. (2011). Towards a sociology of diagnosis: Reflections and opportunities. *Social Science & Medicine, 73*(6), 793–800.

Kannel, W.B., Dawber, T.R., Kagan, A., Revotskie, N., & Stokes, J. (1961). Factors of risk in the development of coronary heart disease – six year follow-up experience: The Framingham study. *Annals of Internal Medicine, 55*(1), 33–50.

Kannel, W.B., Castelli, W.P., McNamara, P.M., & Sorlie, P. (1969). Some factors affecting morbidity and mortality in hypertension: the Framingham study. *Milbank Memorial Fund Quarterly, 47*(3), 116–142.

Kern, S.E. (2012). Why your new cancer biomarker may never work: Recurrent patterns and remarkable diversity in biomarker failures. *Cancer Research, 72*(23), 6097-6101.

Kohli-Laven, N., Bourret, P., Keating, P., & Cambrosio, A. (2011). Cancer clinical trials in the era of genomic signatures: Biomedical innovation, clinical utility, and regulatory-scientific hybrids. *Social Studies of Science, 41*(4), 487–513.

Kreiner, M.J., & Hunt, L.M. (2014). The pursuit of preventive care for chronic illness: turning healthy people into chronic patients. *Sociology of Health & Illness, 36*(6), 870–884.

Labarthe, D.R. (1978). Problems in definition of mild hypertension. *Annals of the New York Academy of Science, 304*(1), 3-14.

Lancet editorial. (1975). Shall I treat mild hypertension? *Lancet, 2*, 695.

Martin, S.A., Boucher, M., Wright, J.M., & Saini, V. (2014). Mild hypertension in people at low risk. *British Medical Journal, 349*, g5432, 1756-1833.

Master, A.M., Dublin L.I., & Marks, H.H. (1951). The normal blood pressure and its clinical implications. *American Statistician, 5*(3), 6-7.

McGann, P.J., & Hutson, D.J. (2011). *Sociology of diagnosis. Advances in Medical Sociology, vol 12.* Bingley, UK: Emerald Publishing.

Moser, M. (2004). More about prehypertension. *Journal of Clinical Hypertension, 6*(9), 483-4.

Moynihan, R. (2012). Preventing overdiagnosis: how to stop harming the healthy. *British Medical Journal,* 344:e3502.

Newall, J. (1967). Problems in the development of a TNM system of staging for cancer of the head and neck. *Bulletin of the New York Academy of Medicine. 43*(11), 959-966.

Norris, R.M., Brandt, P.W.T., Caughey, D.E., Lee, A.J., & Scott, P.J. (1969). A new coronary prognostic index. *Lancet, 1*, 274-8.

Palmer, R.S., Loofbourow, D., & Doering, C.R. (1948). Prognosis in essential hypertension: Eight-year follow-up study of 430 patients on conventional medical treatment. *New England Journal of Medicine, 239*(26), 990-4.

Peters, M.V. (1950). A study of survivals in Hodgkin's disease treated radiologically. *American Journal of Roentgenology and Radium Therapy, 63*, 299–311.

Rabinowitch, I.M., Fowler, A.F., & Bensley, E.H. (1939). Diabetic coma: an investigation of mortalities and report of a severity index for comparative studies. *Annals of Internal Medicine, 12*(9), 1403-28.

Robinson, S.C., & Brucer, M. (1939). Range of normal blood pressure: a statistical and clinical study of 11,383 persons. *Archives of Internal Medicine, 64*(3), 409-444.

Rosenberg, C.E. (2002). The Tyranny of Diagnosis: Specific entities and individual experience. *Milbank Quarterly, 80*(2), 237-260.

Saadatmand, S., Bretveld, R., Siesling, S., & Tilanus-Linthors, M.M.A. (2015). Influence of tumour stage at breast cancer detection on survival in modern times: population based study in 173,797 patients. *British Medical Journal* 351:h4901 doi: 10.1136/bmj.h4901

Sauter, G., & Simon, R. (2002). Predictive molecular pathology. *New England Journal of Medicine, 347*(25), 1995-6.

Schnur, S. (1953). Mortality rates in acute myocardial infarction. II. A proposed method for measuring quantitatively severity of illness on admission to the hospital. *Annals of Internal Medicine, 39*(5), 1018-1025.

Sellers, A.H. (1966). The clinical stage classification of malignant tumours. *Canadian Medical Association Journal, 95*(7), 313-315.

SPRINT Research Group. (2015). A randomized trial of intensive versus standard blood-pressure control. *New England Journal of Medicine, 373*, 2103-16.

Stahlman, M.T., Battersby, E.J., Shepard, F.M., & Blankenship, W.J. (1967). Prognosis in hyaline-membrane disease: Use of a Linear-Discriminant. *New England Journal of Medicine, 276*(6), 303-9.

Timmermans, S., & Stivers, T. (2018). Clinical forecasting: Towards a sociology of prognosis. *Social Science & Medicine, 218*, 13-20.

Truett, J., Cornfield, J., & Kannel, W. (1967). A multivariate analysis of the risk of coronary heart disease in Framingham. *Journal of Chronic Diseases, 20*(7), 511-524.

Vargo-Gogola, T., & Rosen, J.M. (2007). Modelling breast cancer: One size does not fit all. *Nature Reviews Cancer, 7*, 659–672.

Weigelt, B., & Reis-Filho, J.S. (2009). Histological and molecular types of breast cancer: Is there a unifying taxonomy? *Nature Reviews Clinical Oncology, 6*, 718–730.

Will, C.M. (2005). Arguing about the evidence: readers, writers and inscription devices in coronary heart disease risk assessment. *Sociology of Health & Illness, 27*(6), 780–801.

Zieve, L., & Hill, E. (1953). Prognosis in moderate or severe diabetic acidosis. *Archives of Internal Medicine, 92*(1), 63-74.